# Fine-Tuning a Sentiment Analysis Model

# The Problem

- Customer feedback:
  - Surveys
  - Social media
  - Product reviews
- The goal is to create a sentiment analysis model from a pre-trained Hugging Face model

# The Data

- The dataset comes from an online clothing retailer
- There are 10 features and over 20,000 entries

| | Clothing ID | Age | Title | Review Text | Rating | Recommended IND | Positive Feedback Count | Division Name | Department Name | Class Name |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 767 | 33 | NaN | Absolutely wonderful - silky and sexy and comf... | 4 | 1 | 0 | Initmates | Intimate | Intimates |
| **1** | 1080 | 34 | NaN | Love this dress! it's sooo pretty. i happene... | 5 | 1 | 4 | General | Dresses | Dresses |
| **2** | 1077 | 60 | Some major design flaws | I had such high hopes for this dress and reall... | 3 | 0 | 0 | General | Dresses | Dresses |
| **3** | 1049 | 50 | My favorite buy! | I love, love, love this jumpsuit. it's fun, fl... | 5 | 1 | 0 | General Petite | Bottoms | Pants |
| **4** | 847 | 47 | Flattering shirt | This shirt is very flattering to all due to th... | 5 | 1 | 6 | General | Tops | Blouses |

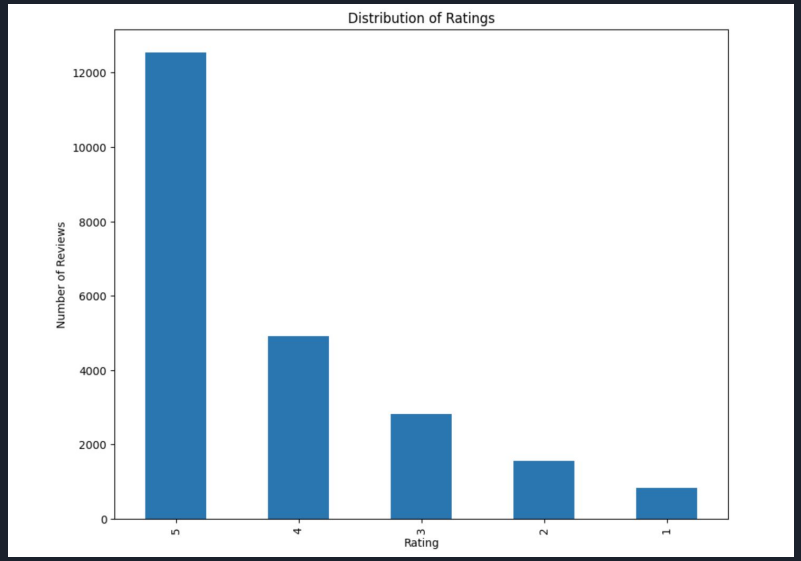
# The Data

- The cleaned dataset contains only 2 columns:
  - Review Text
  - Rating (1-5)

| | Review Text | Rating |
|---|---|---|
| 0 | Absolutely wonderful - silky and sexy and comf... | 4 |
| 1 | Love this dress! it's sooo pretty. i happene... | 5 |
| 2 | I had such high hopes for this dress and reall... | 3 |
| 3 | I love, love, love this jumpsuit. it's fun, fl... | 5 |
| 4 | This shirt is very flattering to all due to th... | 5 |

# The Data

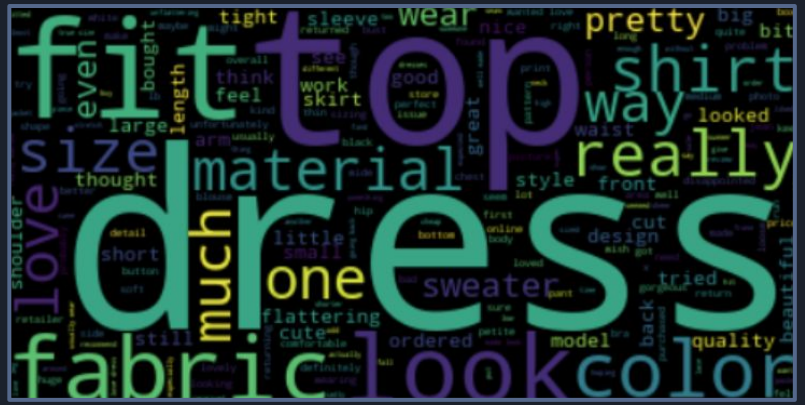
- The possible ratings are 5, 4, 3, 2, and 1
  - 5 is the most positive rating
  - 1 is the most negative rating
- There were more positive ratings of 5 than any other rating



# Visualizing the Data

Rating 5 and 4


Rating 2 and 1



# The Model

- Pre-trained model from Hugging Face
  - Distilbert-base-uncased
  - Smaller and faster version of BERT model

# Training the Model

- The pre-trained model is trained for sentiment analysis using the clothing reviews dataset
- 5 class labels
  - 5, 4, 3, 2, and 1
- Default training arguments were used in training

# Model Metrics

| Evaluation Metrics | Value |
| --- | --- |
| Loss | 0.928 |
| Accuracy | 0.672 |
| F1 Score | 0.668 |
| Runtime | 19.640 |

# Conclusions

- The trained model can be used for sentiment analysis of customer feedback
- Benefits of sentiment analysis:
  - Find out customers feelings toward products
  - Determine areas that need improving
  - Figure out best marketing techniques

# Future Considerations

- Collect more data with a greater variety in rating distribution
- Include different product reviews to expand the scope of the model
- Test different model parameters to improve model performance